

## 2019年度 国立情報学研究所実務研修報告書

2020年6月25日

前田 隼（北海道大学附属図書館）

■研修期間 2020年2月15日～2020年3月15日

■研修先 欧州原子核研究機構（CERN）  
International Digital Curation Conference  
International Open Science Conference  
ETH Zürich Library  
Max Planck Digital Library

### ■研修目的

- 欧州原子核研究機構（CERN）が運営する文献データベース（INSPIRE）を使った、データキュレーションに関する知識と技術の習得
- 欧州におけるオープンアクセス・オープンサイエンスに関する最新動向の把握

### ■研修日程（研修先及び研修期間）

- 2/15-2/16 : 札幌 → 東京 → ダブリン
- 2/17-2/20 : International Digital Curation Conference 2020（アイルランド、ダブリン）
- 2/21-3/6 : CERN 実務研修（スイス、ジュネーブ）  
（NII-前田 Web ミーティング, 2/25 18 JST-10 CET, 3/4 18 JST-10 CET）
- 3/9 : ETH Zürich Library（スイス連邦工科大学チューリヒ校図書館）  
（スイス、チューリッヒ）
- 3/10-3/12 : International Open Science Conference 2020（ドイツ、ベルリン）
- 3/13 : Max Planck Digital Library（ドイツ、ミュンヘン）
- 3/14-3/15 : ミュンヘン → 東京 → 札幌



## ■ 研修成果

### - 要旨

(CERN 研修)

INSPIRE のデータキュレーションを日本国内から行うワークフローとシステムについて INSPIRE Team と共に検討を行った。Inspire 側と WEKO3 側の要求を満たすワークフローとして、WEKO3 のインターフェースを用いて INSPIRE のレコードを直接キュレーションするワークフローを提案した。WEKO3 側では INSPIRE からメタデータを受け取ることで日本国内のリポジトリ・コンテンツの充実につなげる。

(会議出席)

International Digital Curation Conference に出席した。データキュレーションについては図書館職員と研究者の双方についてキュレーションの認識を高めることが必要な段階である。またメタデータ作成のトレーニングも試行されているが、各国で独自にトレーニングツールを開発しており、会議での情報共有と意見交換を行っている段階であった。そのため、欧州として統一した動きはなく、各国ができる範囲で広報と実践に取り組んでいる段階である。メタデータは研究者自身に入力してもらうのがベストだが、実験を行ったところ、10～20 分以上の時間がかかる場合、メタデータ作成に消極的になるとの報告があった。

International Open Science Conference でも

(訪問調査)

日本の機関リポジトリとグリーン・オープンアクセスを見直すため、欧州のオープンアクセスの取り組みについて ETH Zürich Library と Max Planck Digital Library の 2 機関へ訪問調査を実施した。ETHZ Library は機関リポジトリの意義をオープンアクセス手段の一つとして見出している。研究者はゴールド・オープンアクセスによる出版、分野リポジトリによる公開と同様に、選択肢の一つとして機関リポジトリによる公開を選ぶことができる。大学としての研究活動は別の学内データベースに収められており、こちらは学内者の登録が必須となっている。MPDL では OA2020 や ProjektDEAL の概要と進捗説明があり、オープンアクセスという大きな枠でとらえた際に、機関リポジトリによるグリーン・オープンアクセ

スの寄与が小さいことから、現時点で機関リポジトリの果たす役割について再考すべき時期に来ているとの議論があった。日本の機関リポジトリの運営実態を鑑みて、医理工系の欧文雑誌論文のコレクションという位置づけから人文社会系のオープンアクセス手段としてのパラダイムシフトが示唆された。

#### ■研修を終えて

CERN について、一番印象に残っていることは何かときかれれば、それはそこで働くだけでも”Accelerate science”を意識していたことだ。CERN SIS の職員は、なぜ自分がこの仕事をするのか、あるいはしないのかを大きな枠組みの中で明確に理解していた。対して日本の大学図書館を見ていると、惰性で動いている面が散見される。それは一体誰のためなのか、という視点が抜けており、ひどい場合は図書館職員のための仕事のようなものもある。手段が目的化していることも少なくない。

CERN で出張手続きを担当してくれた日本でいうところの庶務担当の Marina はそろそろ定年だろうかという古株だが、現役で CERN guide (一般見学者の施設案内を担当するボランティアで専用の訓練(トレーニングコース)を修了する必要がある)を引き受けており、訪問者に対して電磁気学をかみ砕いて加速器の説明をしたり、検出器 (Detector) を案内したりしている。日本で Lorenz Force を説明しながら自機関の社会的・科学的意義を語れる事務職員はどれほどいるだろうか。Marina に限らず SIS のメンバーの多くは CERN guide として活躍しており、ここでも仕事というものの捉え方を教えられた気がした。

これは何も CERN に感銘を受け、日本に失望したわけではない。重要なのは、CERN 職員は科学への強い貢献意識をもっている点である。このような環境であったため、日本側の事情を説明すると理解はしてもらえもの納得をしてもらうことは難しく、それは容易に想像できたことではあるが、自分が張本人の日本人であることも相まって、正直非常に苦しい状況であった。反面、日本にいてもこれは誰のための仕事だろうと思っていることが多い私にとっては桃源郷を見た気がした。そしてそれは実現すべき、かつ実現可能な理想であるはずである。

リポジトリに限らず、今後日本の科学に貢献できるよう、引き続きやるべきこと(重要課題)に取り組んでいきたい。

末筆になってしまったが、本研修の実現と研修中のサポートに尽力いただいた国立情報学研究所のみなさま、北海道大学のみなさま、そして CERN のみなさまに感謝したい。個人的にも非常に重要な経験になったことは間違いない。今後いい方向に還元していきたい。みなさま、ありがとうございました。

Thank you & Merci.

## 2019 年度 国立情報学研究所実務研修報告書

## 欧州原子核研究機構（CERN）におけるデータキュレーションに関する研修

2020 年 6 月 25 日

前田 隼（北海道大学附属図書館）

## ■背景

2017 年，2018 年に本研修を活用して 2 名の図書館職員が日本から CERN へと派遣された（期間はそれぞれ 1 年，3 カ月）。CERN の運営する高エネルギー物理学（High Energy Physics: HEP）分野のデータベース，INSPIRE の日本人著者情報の充実のためである。具体的な作業は著者の名寄せ（キュレーション）であり，今回この作業を日本国内からリモートで行えるよう，ワークフローとシステムについて現地で INSPIRE Team と共に検討を行った。

なお，論文出版数に応じて各国が分担支出することで HEP 分野の論文をオープンアクセスにする取り組み（SCOAP<sup>3</sup>）が行われている。この分担金の支出を正確に計算するためにも名寄せは重要である。また，SCOAP<sup>3</sup>の論文出版数トップ 5 国で INSPIRE のキュレーションを行っていないのは日本のみであり，ドイツ，アメリカ，イギリス，中国はすでにリモート・キュレーションを確立し実施している。

（先行試行と異なる点）

- INSPIRE がメジャー・アップデートされた（2020 年 3 月）  
（従来のものを legacy，アップデート後のものを labs と呼び分け，本稿では labs によるキュレーション作業を検討の対象にしている）
- リモートによるキュレーション体制の確立を目的としている  
現地に人を派遣してキュレーションを行うのではなく，日本からリモートで作業できる環境を整えることを目的としている

## ■方法

CERN，Scientific Information Service (SIS)，INSPIRE Team のサポートを受けつつ自身も実際にキュレーションを試行したうえで，日本の機関リポジトリの運用事情等も加味しながらワークフローの検討を繰り返した。

2020/2/21-3/6 の CERN 滞在中に，NII-CERN 間の Web ミーティングを 2 回セッションした（2/25 18 JST-10 CET，3/4 18 JST-10 CET）。ワークフローの検討に当たっ

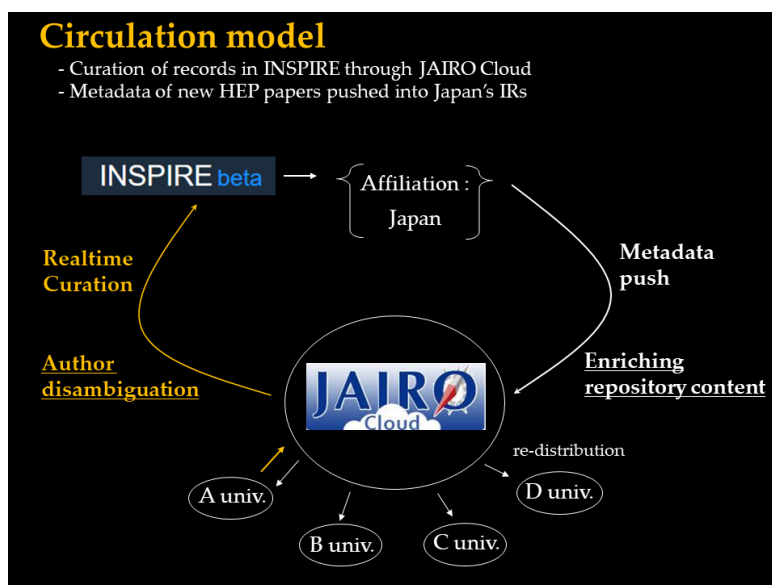
ではこの2回のミーティングで、ワークフロー（案）の提示と議論、フィードバックを受けての改訂を繰り返した。

## ■結果と考察

INSPIRE は arXiv から HEP 分野のプレプリントをハーベストしている。そのうち 60 % はその後査読を経た出版社版として雑誌から出版される。INSPIRE は出版社からこの出版社版もハーベストし、メタデータと本文 PDF を arXiv 取込のデータに対して付加している。このため、論文レコードを基軸とした考え方でキュレーションを行うことは望ましくない。なぜならメタデータを INSPIRE からダウンロードしてキュレーションを行っている最中に出版社版のメタデータが付加された場合、アップロードしたメタデータで上書きすることは出版社版メタデータとの衝突を引き起こすためである。そのため、INSPIRE 側では常にリアルタイム・キュレーションが要望されており、各国のリモート・キュレーションも直接キュレーターたちが INSPIRE にログインしてリアルタイムにキュレーションを実施している。

一方、日本側では INSPIRE のみを対象としたキュレーターの養成・チームの編成が現実的ではないため、既存の機関リポジトリ運用ワークフローの中にこのキュレーションを取り込むのが一手である。INSPIRE へのログインは ORCID を用いるが、日本では担当・部署異動が常であるため、このログイン方法では毎年 INSPIRE 側も日本側も手続きが煩雑である。

このような 2 者の異なる要求の中で、本稿が提示したのは「循環モデル (Circulation model)」である。これは CERN の衝突器である LHC (Large Hadron Collider) をモデルにしたもので、INSPIRE を起点として、日本のリポジトリへメタデータをプッシュし、日本のリポジトリからは著者の漢字名 (native names) を INSPIRE へ返すというモデルである。



## ■今後

### - スモール・スタート

2020年6月のNII関係者とのミーティングで、メタデータのダウン・アップロードの方法を用いてINSPIRE-WEKO3間の循環モデルを試行する方向で調整に入った。これはINSPIREのレコードを直接キュレーションせず、データの上書きによってnative namesを提供するものだが、第一歩となることを期待したい。

### - ORCIDの新たな側面

循環モデル実現の際に問題となるのは、リアルタイムでのキュレーションをいかに実現するかということである。ORCIDをキュレーション従事者が取得しINSPIRE側でオーソライズしてもらうという流れは現状では煩雑だが、ORCIDを業務用に取得することが日本の図書館職員の間で普及していけば、INSPIREに限らずキュレーションを行う際のオーソライズIDとして活用できるのではないだろうか。

### - HEPのその先へ 循環モデルの他分野への応用

本稿ではINSPIREを対象にキュレーションの検討を行っているが、本来これはHEP分野のみに限定された話ではない。つまり、循環モデルの適用例のひとつとしてHEPがあり、連携先がINSPIREであるというのが、中立的な考え方だろう。だとすると、この循環モデルはリモート・キュレーションモデルとして他分野への応用可能性もあるはずで、このワークフローと運用体制が整備されれば、広範囲の学術分野に貢献できることになる。このことは最重要ポイントであり、常に覚えておきたい。

2019 年度 国立情報学研究所実務研修報告書

International Digital Curation Conference 2020

会議参加報告

2020 年 6 月 25 日

前田 隼（北海道大学附属図書館）

会議名： International Digital Curation Conference 2020

期間： 2020 年 2 月 16 日～2 月 21 日

会場： Croke Park, Dublin, Ireland



(主なトピック)

**FAIRsFAIR**

- 欧州各国では FAIRsFAIR を実行すべく、図書館員向け、研究者向けのトレーニングツールを開発・試行中である。データの保存に関しては、誰と何をどこまで共有するか、どこに保存するかが課題となっている

**Data Curation**

- データ保存の際のメタデータをどのように作るか（データキュレーション）  
研究者自身に入力してもらうのがベストだが、実験を行ったところ、10～20 分以上の時間がかかる場合、メタデータ作成に消極的になる傾向が明らかになった
- そこで事前に選択肢を準備してプルダウン式の回答フォームを作り省力化を図ったが、今度は正確性との兼ね合いが課題となった

- ・ FAIR 原則を履行してもらいたいですがペナルティを科したり、厳格に管理したりすることは図書館側としてはやりたくない
- ・ PhD 学生への広報と教育に力を入れるのがベストではないか

#### **RDM/DMP**

- ・ ノルウェーとアメリカの調査結果では、現時点でデータ管理計画（DMP）と質と助成金採否の間には明確な関連性が見られなかった
- ・ DMP の目的はそれが DMP として受け付けられることであり、それ自体に手間暇がかかっている本末転倒である。本当のゴールは科学的成果である。
- ・ 人文社会学分野のデータ管理には慎重な姿勢が求められる。なぜなら個人情報やプライバシーなど法的リスクを負うものがあるためである

#### **Open Science**

- ・ EU 調べでは、85 %の研究が FAIR という観点で要件を満たしておらず、これは 10.2 億ユーロの損失に値する
- ・ PhD 学生への広報と教育によって変えていくのが一手だろう



2019 年度 国立情報学研究所実務研修報告書

International Open Science Conference 2020

会議参加報告

会議名： International Open Science Conference 2020

期間： 2020 年 3 月 11 日～3 月 12 日 (2 日目は COVID-19 対策のため中止)

会場： H4 Hotel, Berlin, Germany



**Open Science**

- ・ オープンサイエンスの推進にあたって、研究者の理解を得るにはどうしたらよいか  
→インセンティブや取り組みが評価されることが必要
- ・ オープンサイエンスについて研究者への認知が低いため、広報やサポートが必要
- ・ データの公開については法的な問題や、DMP との関係、ひいては助成金採択との関係など考慮すべき裾野が広い
- ・ 研究者からの反発から、オープンサイエンスの推進には 4 つの壁があると考えられる
  - 認知度 (現状オープンサイエンスの取り組みはあまり認知されていない)
  - 関連性 (自分の分野との関連性がよくわからない)
  - 効率性 (研究の効率を下げるのではないか)
  - 持続性 (持続的に行うだけのツールや支援があるのか)

**Open data**

- ・ 質的データ (アンケートやインタビュー) を公開できるのか

- ・オープンデータによるシチズン・サイエンスの促進が見込める一方、データや結果の信頼性をどう担保するのか

2<sup>nd</sup> day

---Closed---

2019 年度 国立情報学研究所実務研修報告書

ETH Zürich Library & Max Planck Digital Library

訪問調査報告

2020 年 6 月 25 日

前田 隼 (北海道大学附属図書館)

(訪問調査 1)

訪問先：ETH Zürich Library (スイス連邦工科大学チューリヒ校図書館)

対応者：Dr. Rainer Rees-Mertins

Dr. Matthias Töwe

Barbara Hirschmann

トピック：機関リポジトリとデータキュレーションの実務について

(訪問調査 2)

訪問先：Max Planck Digital Library

対応者：Dr. Ralf Schimmer, Head of Information

Colleen Campbell, Open Access 2020 Initiative

トピック：ヨーロッパにおけるオープンアクセスの取り組みについて

**IR : Leave it as a choice for OA**

**ETH Zürich Library**

- Scientific activity documenting system mandatory for all faculty members
- IR as a choice for OA  
(No PDF request e-mail to authors)

**Max Planck Digital Library**

- Green OA cannot be a solution of open access: OA2020 is the best way
- Time to think the purpose/value of IRs



しかし現在、すでに状況は大きく変わってきた。OA という流れの中でグリーン OA がメインストリームとならない今、ゴールド OA と並行して OA 出版手段として位置づける必要があるのではないだろうか。

同時に、機関リポジトリからの視点でコンテンツを考えるのではなく、むしろコンテンツ側、つまり学術コミュニティ側からみた機関リポジトリの使い方を考える必要があるだろう。

以上の背景から、日本の機関リポジトリの方向性について以下の 2 点を提案したい。

(医理工系分野)

- ・機関リポジトリを OA 手段の一つとして位置づける。本文提供依頼は、そのコンテンツを機関リポジトリで公開することではじめて OA になる場合に行う。

(人文社会学分野)

- ・紀要に加えて、雑誌論文、研究データの格納場所として位置づける。機関リポジトリの運用も人文社会学分野のサポートにシフトする。